



TITLE:

複雑ネットワークにおけるコミュニティ構造の中核部分:抽出法の開発と応用(経済物理学とその周辺,統計数理研究所研究会共同研究集会,経済物理学2009-ミクロとマクロの架け橋-,京都大学基礎物理学研究所2009年度前期研究会,研究会報告)

AUTHOR(S):

飯野, 隆史; 家富, 洋

CITATION:

飯野, 隆史 ...[et al]. 複雑ネットワークにおけるコミュニティ構造の中核部分:抽出法の開発と応用(経済物理学とその周辺,統計数理研究所研究会共同研究集会,経済物理学2009-ミクロとマクロの架け橋-,京都大学基礎物理学研究所2009年度前期研究会,研究会報告). 物性研究 2010, 93(5): 670-671

ISSUE DATE:

2010-02-05

URL:

<http://hdl.handle.net/2433/169226>

RIGHT:

複雑ネットワークにおけるコミュニティ構造の中核部分 — 抽出法の開発と応用 —

新潟大院自然, 新潟大理^A 飯野 隆史¹, 家富 洋^A

1 目的と対象

コミュニティとはネットワークにおける相対的にリンクが集中している部分のことであり、複雑ネットワークを特徴づける大事な要素である。従来のコミュニティ抽出方法 [1] では、コミュニティに強く属するノードと弱く属するノードが一括りで抽出されてしまう。そこで本研究では、コミュニティ抽出に温度の概念を取り入れ、コミュニティ内構造の安定性を測ることで、中核部分（特に強くリンクしている集団）を取り出すことを目的とする。対象となる企業間取引ネットワーク²は、企業をノード、企業間の取引関係をリンクとしたネットワークである。ノード数約 80 万、リンク数約 320 万の大規模なネットワークとなっている。本研究と同じデータを用いて、製造業に限った取引ネットワークについて既にコミュニティ解析が行われている [2]。

2 手法

Newman 等によるコミュニティ抽出では、モジュラリティ (modularity) Q と呼ばれる指標を用いる (Q の具体的な表式は文献 [1] を参照)。同じコミュニティに属するノード内に多くのリンクがあれば、 Q の値が大きくなる。指標 Q を最大とする分割が最適なコミュニティ構造を与える。

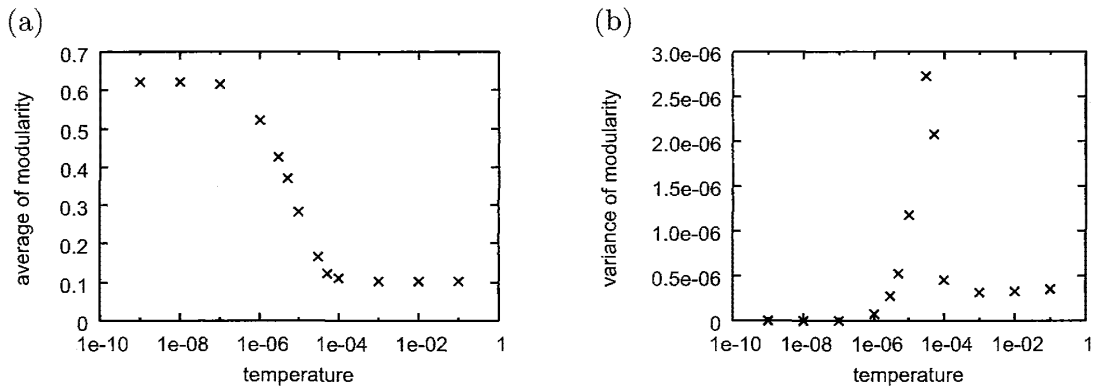
本研究では、温度の概念を取り入れ、コミュニティの熱平衡状態のアンサンブルをとる。大規模なネットワークのコミュニティ抽出においては、greedy アルゴリズム [3] がよく用いられるが、貪欲に最適値を求めるだけでは有限温度のアンサンブルは得られない。そこで、メトロポリス法によるコミュニティ抽出を行う。1 ステップごとにノードをランダムに選択し、コミュニティの更新を検討する。選ばれたノード v に隣接するノードの数を k_v とすると、ノード v は $1/(k_v + 1)$ の確率で、隣接ノードのコミュニティの一つに所属変更する。どの隣接ノードも選ばれない場合も $1/(k_v + 1)$ の確率で出現し、そのときはネットワーク上に無い新しいコミュニティを導入する。変更後と変更前のモジュラリティの差分 ΔQ が、 $\Delta Q > 0$ のときは変更を受け入れ、 $\Delta Q < 0$ のときは

$$P(\Delta Q, T) = \exp \left\{ \frac{\Delta Q}{T} \right\} \quad (1)$$

の確率で変更を受け入れる。ここで、 T は温度に対応したパラメータであり、 T が高くなると、低いモジュラリティ状態でも受け入れやすくなる。

¹E-mail: iino@ad.sc.niigata-u.ac.jp

²東京商工リサーチ (TSR) のデータを、独立行政法人経済産業研究所 (RIETI) から提供を受けた。

図 1: モジュラリティ Q の温度変化: (a) 平均値, (b) 分散

3 結果と考察

企業間取引ネットワークについて、各温度でノード数 $\times 100$ ステップ行い、それを繰り返してコミュニティ構造のアンサンブルを 400 回とった。各温度でのモジュラリティの平均、分散を図 1 に示す。温度が上がるほど構造が乱されてモジュラリティの平均値は小さくなり、 $T = 10^{-5}$ 付近で平均値が大きく変化する。一方、分散の温度変化を見てみると、 $T = 10^{-5}$ 付近のところでピークが見られる。モジュラリティにおける“比熱” dQ/dT と揺らぎが対応していることを示している。

400 回分のコミュニティの積集合をとることで、結びつきの強い中核部分を抽出した。表 1 は、最大のコミュニティから分かれた各温度の中核部分に含まれる企業の地域・業種の割合をまとめたものである。温度が上がるほど、中核部分のサイズは小さくなる。北海道企業の割合はネットワーク全体に比較してもともと高いが、温度が上がるにつれさらに高くなる。建設材料卸売業については $T = 0$ では目立たなかったが、 $T = 10^{-6}$ の中核部分では割合が高くなり、強い結びつきを持つことが分かる。一方、建設業の割合は $T = 0$ では非常に高かったが、中核部分をとることで割合が低下し、建設材料卸売業に比較して結びつきの弱い企業も含まれていることが分かる。他のコミュニティについても同様に、繋がり弱い部分を削ぎ落とすことによって、コミュニティの中で強い繋がりを持つ重要な地域・業種を抽出することができる。

	$T = 0$	$T = 10^{-8}$	$T = 10^{-7}$	$T = 10^{-6}$	ネットワーク全体
サイズ	195,461	194,442	104,664	6,824	780,544
北海道 (%)	14.8	14.8	18.9	27.3	6.3
建設業 (%)	71.6	71.7	72.2	51.5	32.7
建設材料卸売業 (%)	6.0	6.0	7.6	20.8	4.1

表 1: 最大コミュニティから分かれた中核部分に含まれる企業属性の割合

参考文献

- [1] M. E. J. Newman, Phys. Rev. E **69** (2004), 066133; PNAS **103** (2006), 8577.
- [2] Y. Fujiwara and H. Aoyama, 0806.4280[e-print arXiv].
- [3] A. Clauset, M. E. J. Newman and C. Moore, Phys. Rev. E **70** (2004), 066111.